

Sent to Lum. Bat.
01/07/
545834

Pyrosequencing

A bioluminometric method of DNA sequencing

Mostafa Ronaghi¹, Nader Pourmand¹ and Viktor Stolc²

Address

¹Stanford Genome Technology Center, Stanford University,
Palo Alto, CA, USA. ²NASA Ames Research Center, Moffett
Field, CA, USA.

Correspondence can be addressed:

Mostafa Ronaghi

Genome Technology Center

Stanford University

855 California Avenue

Palo Alto, CA 94304

Phone: (650) 812-1971

Fax: (650) 812-1975

e-mail: mostafa@stanford.edu

ABSTRACT

Pyrosequencing is a bioluminometric single-tube DNA sequencing method that takes advantage of co-operativity between four enzymes to monitor DNA synthesis. In this sequencing-by-synthesis method, a cascade of enzymatic reactions yields detectable light, which is proportional to incorporated nucleotides. Pyrosequencing has the advantages of accuracy, flexibility and parallel processing. It can be easily automated. Furthermore, the technique dispenses with the need for labeled primers, labeled nucleotides and gel-electrophoresis. In this chapter, the use of this technique for different applications is discussed.

INTRODUCTION

The development of DNA sequence determination techniques with enhanced speed, sensitivity and throughput is of the utmost importance for the study of biological systems. Conventional DNA sequencing relies on the elegant principle of the dideoxy chain termination technique first described more than two decades ago (Sanger et al. 1977). This multi step principle has gone through major improvements during the years to make it a robust technique that has been used for sequencing of several different bacterial, archeal and eucaryotic genomes (www.ncbi.nlm.nih.gov, and www.tigr.org). However, this technique faces limitations in both throughput and cost for most future applications. Many research groups around the world have put efforts into the development of alternative principles for DNA sequencing. Three methods, which hold great promise are sequencing by hybridization (Bains & Smith 1988; Drmanac et al. 1989; Khrapko et al. 1989; Southern 1989), parallel signature sequencing based on ligation and cleavage (Brenner et al. 2000) and Pyrosequencing (Ronaghi et al. 1996; Ronaghi et al. 1998a).

Pyrosequencing is based on the detection of released pyrophosphate (PPi) during DNA synthesis. In a cascade of

enzymatic reactions, visible light is generated which is proportional to the number of incorporated nucleotides. The cascade starts with a DNA polymerization reaction in which inorganic pyrophosphate (PPi) is released as a result of nucleotide incorporation by polymerase (Figure 1). The released PPi is subsequently converted to ATP by ATP sulfurylase. The synthesized ATP provides the energy to luciferase to oxidize luciferin and generate light. Unincorporated deoxy nucleotides and ATP are degraded by the enzyme apyrase. Since the identity of the added nucleotide is known at each step of the light signal generation, the sequence of the template can be determined. Standard Pyrosequencing uses the Klenow fragment of *E. coli* DNA polymerase I, which is a relatively slow polymerase (Benkovic & Cameron 1995). The ATP sulfurylase used in Pyrosequencing is a recombinant version from the yeast *Saccharomyces cerevisiae* (Karamohamed et al. 1999). The luciferase is from the American firefly *Photinus pyralis* (DeLuca & McElroy 1984) and the nucleotide-degrading enzyme, apyrase, is from potato tubers, *Solanum tuberosum* (Handa & Guidotti 1996). The polymerization and detection reactions takes place within 3-4 seconds at room temperature. Efficient enzymatic nucleotide removal takes approximately 20-30 seconds. Recent improvements in the

Pyrosequencing reaction include the use of dATP α S instead of dATP in polymerization (Ronaghi et al. 1996), inclusion of apyrase (Ronaghi et al. 1998a), and inclusion of single-stranded DNA-binding protein (Ronaghi 2000). These improvements have resulted in a read-length of more than 100 nucleotides. In this chapter, we describe the use of this technique for accurate analysis of single nucleotide polymorphisms (SNPs), tag sequencing, and microbial typing.

MATERIALS AND METHODS

In vitro amplification and template preparation

PCR was performed according to the standard protocols using one 5' biotinylated primer. The products were immobilized onto streptavidin-coated paramagnetic beads according to the suppliers recommendations (Dynal A.S., Oslo, Norway). Single-stranded DNA was obtained by removing the supernatant after incubating the immobilized PCR product in 0.10 M NaOH for 3 min. Sequencing primer was hybridized to the immobilized single-stranded DNA strand in 10 mM Tris-acetate pH 7.5, and 20 mM magnesium acetate. For multiplex Pyrosequencing, several sequencing primers were hybridized to a single template simultaneously.

Pyrosequencing

Pyrosequencing™ was performed at room temperature in a volume of 50 µl on a PSQ™ 96 system (www.pyrosequencing.com). DNA template with hybridized primer was added to the Pyrosequencing reaction mixture containing: 10 U exonuclease-deficient Klenow DNA polymerase (Amersham Pharmacia Biotech, Uppsala, Sweden), 40 mU apyrase (Sigma Chemical Co., USA), 100 ng purified

luciferase (BioThema, Dalarö, Sweden), 15 mU of recombinant produced ATP sulfurylase (Karamohamed et al. 1999), 0.1 M Tris-acetate (pH 7.75), 0.5 mM EDTA, 5 mM magnesium acetate, 0.1% bovine serum albumin, 1 mM dithiothreitol, 5 μ M adenosine 5'-phosphosulfate (APS), 0.4 mg/ml polyvinylpyrrolidone (360 000), and 100 μ g/ml D-luciferin (BioThema). For determination of long reads of DNA sequence (<100 nucleotides) and for Pyrosequencing of difficult templates with high GC content and potential for the formation of secondary structure, 0.5 μ g SSB (Amersham Pharmacia Biotech) was added to the DNA template. For multiplex Pyrosequencing, three sequencing primers were hybridized on a single-stranded DNA template. The sequencing procedure was carried out by stepwise elongation of the primer-strand upon sequential addition of the different deoxynucleoside triphosphates (Amersham Pharmacia Biotech) and simultaneous degradation of nucleotides by apyrase. The output of light resulting from nucleotide incorporation was detected by a CCD-camera (Charge-coupled device-camera) in the PSQ[™] 96 system.

RESULTS AND DISCUSSION

Pyrosequencing is emerging as a widely applicable technology for the detailed characterization of nucleic acids. DNA template can be sequenced in real-time enabling very fast analysis. Using one pmol of DNA, 6×10^{11} ATP molecules can be obtained which, in turn, generate more than 6×10^9 photons at a wavelength of 560 nanometers. This amount of light is easily detected by a photodiode, photomultiplier tube, or a CCD-camera.

An automated version of a Pyrosequencing machine (PSQ™ 96 system) was recently developed (www.pyrosequencing.com). The automated version of Pyrosequencing uses a disposable inkjet cartridge for precise delivery of small volume (200 nl) of six different reagents into a temperature controlled microtiter plate (Figure 2). The microtiter plate is under continuous agitation to increase the rate of reagent mixing. A lens array is used to efficiently focus the generated luminescence from each individual well of the microtiter plate onto the chip of a CCD-camera. Nucleotides are dispensed into alternating wells with a delay to minimize mixing of generated light between different wells. A cooled CCD-camera images the plate every second to follow

the exact process of the Pyrosequencing reaction. Data acquisition modules and an interface for PC-connection are used in this instrument. Pyrosequencing software running in Microsoft Windows enables individual control of the dispensing order for each well in the PSQ™ 96 system.

Prior to initiating Pyrosequencing, the reagents and each of the four nucleotides are loaded into the inkjet cartridge that is mounted in the instrument. A microtiter plate containing DNA template with hybridized primer is placed into the Pyrosequencing machine, and after the enzymes and substrate has been delivered by the ink-jet, nucleotides are added to the solution according to the specified order. The Pyrosequencing signals in a pyrogram (Figure 3, 4, and 5) show high quality sequence data with high signal-to-noise ratio. The height of the peaks is proportional to the number of incorporated nucleotides. A 384-wells microtiter plate-based high throughput version of this machine is also under development, which will allow the analysis of up to 50,000 SNPs per 8 hours and reduce the cost for genotyping below 20 cents per sample (www.pyrosequencing.com).

In addition to Pyrosequencing, the PSQ™ 96 System is a high-precision 96 well luminometer that can be used for any enzyme-screening assay that generates a light signal. For example, the PSQ™ 96 System can be used for any high-throughput ELISA-based immunoassays. It can also be used for screening of all enzyme activities found in yeast cells, by using genome-wide epitope-tagged yeast strains, each containing a different yeast open reading frame (ORF) fused to glutathione *S*-transferase (GST) (Martzen, M.R. et al. 1999, Viktor Stolc, under investigation). The PSQ™ 96 system has already been used for a number of applications (Ronaghi 2001). Three main applications of this machine are presented and discussed bellow.

Pyrosequencing for SNP analysis

For analysis of SNPs by Pyrosequencing, the 3'-end of a primer is designed to hybridize one or a few bases 5' adjacent to the polymorphic position. In a single tube, all the different variations can be determined as the region is sequenced. A striking feature of pyrogram readouts for SNP analysis is the clear distinction between the various genotypes of each allele combination. Thus, homozygous or heterozygous alleles result in a specific pattern (Ahmadian et al. 2000a; Alderborn et al. 2000; Ekstrom et al. 2000;

Nordstrom et al. 2000b). This feature makes genotyping extremely accurate and easy. Relative standard deviation values for the ratio between key peaks of the respective SNPs and reference counterparts are 0.1 or lower (Alderborn et al. 2000). Simple manual comparison of predicted SNP patterns and the raw data obtained from the PSQ™ 96 system can score an SNP, especially as no editing is needed (Figure 3). Because specific patterns can be readily achieved for the individual SNPs, it will also be possible to automatically score the allelic status by pattern recognition software. In a study based on results from three different laboratories, 26 different SNPs and more than 1600 DNA samples were analyzed. The algorithm classified the data from 94% of the samples as good or medium quality and 99.4% of these were automatically assigned the expected genotypes. The primary source of any low data quality was insufficient signal-to-noise ratio, typically caused by low efficiency in PCR amplification. As Pyrosequencing signals are very quantitative, it is possible to use this strategy to examine allelic frequency in large population. Furthermore, Pyrosequencing enables determination of the phase of SNPs when they are in the vicinity of each other allowing the detection of haplotypes

(Ahmadian et al. 2000b). This system allows more than 5000 samples to be analyzed in 8 hours.

Pyrosequencing for tag sequencing

Tag sequencing can be used for de novo sequencing and re-sequencing. This includes applications such as partial cDNA sequencing (Nordstrom et al. submitted, Ronaghi et al. 1998b), microbial typing (Gharizadeh et al., in preparation), re-sequencing of disease gene (Garcia et al. 2000), and identification of oligonucleotide bar-code sequences (http://www-sequence.stanford.edu/group/yeast_deletion_project/deletion_s3.html, Stolc, et al. under investigation). For partial cDNA analysis, theoretically, eight or nine nucleotides in a row should define a unique sequence for every gene in the human genome. However, it has been found that in order to uniquely identify a gene from a complex organism such as human, a longer sequence of DNA is needed (Nordstrom et al. submitted). In a pilot study, it was found that 98% of genes could be uniquely identified by sequencing a length of 30 nucleotides. Pyrosequencing was used to sequence this length for gene identification from a human cDNA library and the results were in complete agreement with longer sequence data obtained by Sanger DNA sequencing.

Pyrosequencing offers high throughput analysis of cDNA libraries since 96 samples can be analyzed in less than one hour. Like Sanger DNA sequencing, Pyrosequencing also has the advantage of library screening, since the original cDNA clone is directly available for further analysis (Figure 4). Pyrosequencing has also been used to sequence 10-50 nucleotides for semi-conserved genes for viral, bacterial and fungal typing (Gharizadeh et al. 2001)

Multiplex Pyrosequencing

Pyrosequencing takes advantage of enzymatic reactions to determine the sequence of the DNA. Proportional signals obtained in a pyrogram represent the number of nucleotides incorporated by DNA polymerase. By using more than one sequencing primer in a Pyrosequencing reaction, a sequence-specific fingerprint can be obtained in which the height of each peak determines the number of incorporated nucleotides. Expected pattern for one of the subtype of hepatitis C virus and the raw data from Pyrosequencing presenting the subtype is demonstrated in Figure 5.

Challenges in Pyrosequencing

An inherent problem with the described method is *de novo* sequencing of polymorphic regions in heterozygous DNA material. In most cases, it will be possible to detect the

polymorphism. If the polymorphism is a substitution, it will be possible to obtain a synchronized extension after the substituted nucleotide. If the polymorphism is a deletion or insertion of the same kind as the adjacent nucleotide on the DNA template, the sequence after the polymorphism will be synchronized. However, if the polymorphism is a deletion or insertion of another type, the sequencing reaction can become out of phase, making the interpretation of the subsequent sequence difficult. If the polymorphism is known, it is always possible to use programmed nucleotide delivery to keep the extension of different alleles synchronized after the polymorphic region. It is also possible to use a bi-directional approach (Ronaghi et al. 1999) whereby the complementary strand is sequenced in order to decipher the sequence flanking the polymorphism.

Another inherent problem is the difficulty in determining the number of incorporated nucleotides in homopolymeric regions, due to the non-linear light response following incorporation of more than 5-6 identical nucleotides. The polymerization efficiency over homopolymeric regions has been investigated and the results indicate that it is possible to incorporate up to ten identical adjacent

nucleotides in the presence of apyrase (Ronaghi 2000). However, to elucidate the correct number of incorporated nucleotides, it may be necessary to use specific software algorithms that integrate the signals. For re-sequencing, it is possible to add the same nucleotide twice for a homopolymeric region to ensure complete polymerization, and for the tag software to automatically sum up the signals representing the correct number of incorporated nucleotides.

Conclusion

Pyrosequencing technology is relatively new and there is much room for developments in both chemistry and in instrumentation. The technology is already time- and cost-competitive when compared to the existing sequencing methods. The current cost is 69 cents per sample using standard Pyrosequencing. With the 384-format machine, the cost will be reduced below 20 cents. Work is underway to further improve the chemistry, to measure the sequencing efficiency at elevated temperatures and to run the reaction in miniaturized formats. The advantage of Pyrosequencing in miniaturized formats may lie in the ease with which large numbers of high-density arrays can be manufactured and the future integration of sample preparation with these devices. Success in miniaturization of this technique into

high density microtiter plates, microarrays or microfluidics (Ronaghi 1998c) will reduce the cost and increase the throughput by one to two orders of magnitude, a crucial step for large-scale genetic testing.

ACKNOWLEDGEMENTS

The authors are supported by NIH grant and The Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation. We would like to thank Ronald Davis for valuable discussion. The authors also acknowledge the Genome Research for the permission to republish Figure 2 and some parts of the text of this chapter.

REFERENCES

- Ahmadian, A., Gharizadeh, B., Gustafsson, A., C., Sterky, F., Nyren, P., Uhlen, M. and Lundeberg, J. 2000a. Single-nucleotide polymorphism analysis by Pyrosequencing. *Anal. Biochem.* 280: 103-110.
- Ahmadian, A., Lundeberg, J., Nyren, P., Uhlen, M. and Ronaghi, M. 2000b. Analysis of the p53 tumor suppressor gene by pyrosequencing. *BioTechniques* 28: 140-144.
- Alderborn, A., Kristofferson, A. and Hammerling, U. 2000. Determination of single nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. *Genome Res.* 10: 1249-1258.
- Bains, W. and Smith, G. C. 1988. A novel method for nucleic acid sequence determination. *J. Theoret. Biol.* 135: 303-307.
- Benkovic, S. J. and Cameron, C. E. 1995. Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment of *Escherichia coli* DNA polymerase I. *Methods Enzym.* 262: 257-269.

Brenner, S., Williams, S. R., Vermaas, E. H., Storck, T., Moon, K., McCollum, C., Mao, J. I., Luo, S., Kirchner, J. J., Eletr, S., DuBridge, R. B., Burcham, T. and Albrecht, G. 2000. In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA.* 97: 1665-1670.

DeLuca, M. and McElroy, W. D. 1984. Two kinetically distinguishable ATP sites in firefly luciferase. *Biochemical & Biophysical Research Communications* 123: 764-770.

Drmanac, R., Labat, I., Brukner, I. and Crkvenjakov, R. 1989. Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics* 4: 114-128.

Eckersten, A., Örlfors, A., E., Ellström, C., Erickson, A., Löfman, E., Eriksson, A., Eriksson, S., Jorsback, A., Tooke, N., Derand, H., and et al. 2000. High-throughput SNP scoring in a disposable microfabricated CD device. *Proceeding of the Micro Total Analysis Systems, Kluwer Academic Publishers:* 521-524.

Ekstrom, B., Alderborn, A. and Hammerling, U. 2000. Pyrosequencing for SNPs. *Progress in biomedical optics* 1: 134-139.

Garcia, A., C., Ahamdian, A., Gharizadeh, B., Lundeberg, J., Ronaghi, M. and Nyren, P. 2000. Mutation detection by Pyrosequencing: sequencing of exons 5 to 8 of the p53 tumour supressor gene. *Gene* 253:249-257.

Handa, M., and Guidotti, G. 1996 Purification and cloning of a soluble ATP-diphosphohydrolase (apyrase) from potato tubers (*Solanum tuberosum*). *Biochem. Biophys Res. Commun.* 218: 916-923.

Karamohamed, S., Nilsson, J., Nourizad, K., Ronaghi, M., Pettersson, B. and Nyren, P. 1999. Production, purification, and luminometric analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*. *Prot. Exp. & Purif.* 15: 381-388.

Khrapko, K. R., Lysov Yu, P., Khorlyn, A. A., Shick, V. V., Florentiev, V. L. and Mirzabekov, A. D. 1989. An

oligonucleotide hybridization approach to DNA sequencing.
FEBS Lett. 256: 118-122.

Nordstrom, T., Nourizad, K., Ronaghi, M. and Nyren, P.
2000a. Methods enabling Pyrosequencing on double-stranded
DNA. *Anal. Biochem.* 282: 186-193

Nordstrom, T., Gharizadeh, B., Pourmand, N., Nyren, P., and
Ronaghi, M 2001. Methods enabling fast partial sequencing
of cDNA clones. *Anal. Biochem.* Submitted.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. and
Nyren, P. 1996. Real-time DNA sequencing using detection of
pyrophosphate release. *Anal. Biochem.* 242: 84-89.

Ronaghi, M., Uhlen, M. and Nyren, P. 1998a. A sequencing
method based on real-time pyrophosphate. *Science* 281: 363-
365.

Ronaghi, M., Pettersson, B., Uhlen, M. and Nyren, P. 1998b.
PCR-introduced loop structure as primer in DNA sequencing.
BioTechniques 25: 876-884.

Ronaghi, M. 1998c. Pyrosequencing: a tool for sequence-based DNA analysis. *Doctoral thesis*, ISBN 91-7170-297-0, The Royal Institute of Technology, Stockholm, Sweden.

Ronaghi, M., Nygren, M., Lundeberg, J. and Nyren, P. 1999. Analyses of secondary structures in DNA by pyrosequencing. *Anal. Biochem.* 267: 65-71.

Ronaghi, M. 2000. Improved performance of Pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* 286: 282-288

Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* In press.

Sanger, F., Nicklen, S. and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74: 5463-5467.

Southern, E. M. 1989. Analysing polynucleotide sequences. *Patent WO/10977.*

FIGURE LEGENDS

Figure 1. The general principle behind different Pyrosequencing reaction systems. A polymerase catalyzes incorporation of nucleotide(s) into a nucleic acid chain. As a result of the incorporation, a PPi molecule(s) is released and subsequently converted to ATP, by ATP sulfurylase. Light is produced in the luciferase reaction during which a luciferin molecule is oxidized. The left over of nucleotides and generated ATP are degraded by the enzyme apyrase. Nucleotides are added one by one within a minute interval and the target sequence is determined.

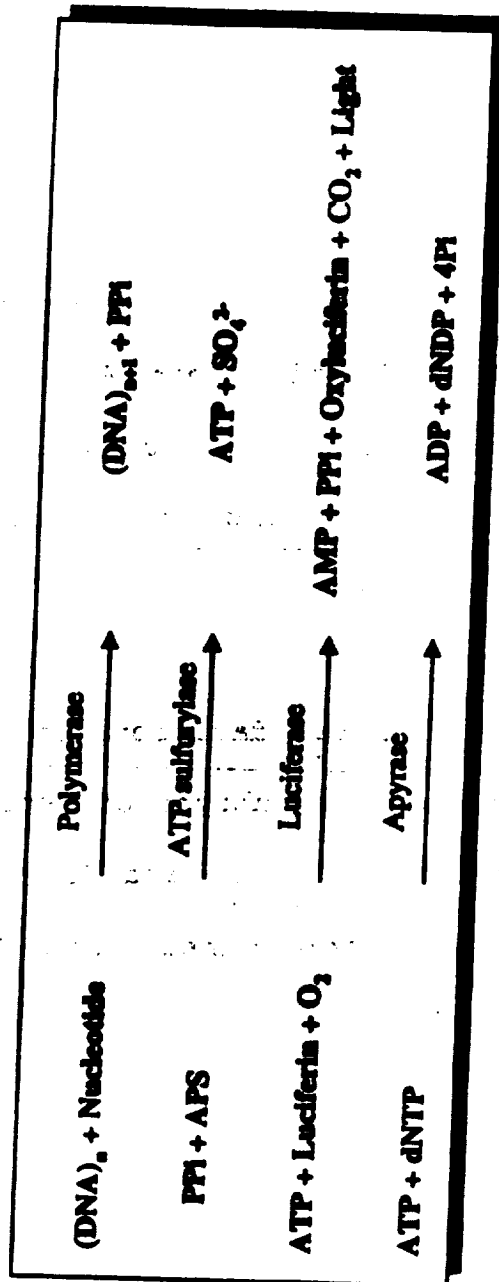
Figure 2. Schematic drawing of the automated system for Pyrosequencing. Four dispensers move on an X-Y robotics arm over the microtiter plate and add four different nucleotides, according to the pre-specified order. The microtiter plate is agitates continuously to mix the added nucleotide. Generated light is directed to the CCD-camera using a lens-array located exactly below the microtiter plate. Reprinted by kind permission of Genome Research.

Figure 3. Pyrogram of the raw data obtained from Pyrosequencing of two different kinds of homozygous alleles

(a and c) and a heterozygote DNA template (b). The order of nucleotide additions is indicated below the pyrogram, and the obtained sequence is indicated above the pyrogram. The height of the peaks determines the number of incorporated nucleotides. The time between addition of each nucleotide is one minute.

Figure 4. Pyrogram of the raw data obtained from *de novo* sequencing of a 230-nucleotide-long PCR product using a PSQ™ 96 system. The order of the nucleotide additions is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram.

Figure 5. Pyrogram of the raw data obtained from multiplex Pyrosequencing of one of the subtype of hepatitis virus C. Three sequencing primers (P1, P2, and P3) were hybridized to a single template (top panel) and Pyrosequencing resulted in a unique fingerprint for a specific genotype. The order of the nucleotide additions is indicated below the pyrogram and the obtained sequence is indicated above the pyrogram. The expected theoretical pattern (left) and the raw data (right) are shown. See Materials and Methods for more information.

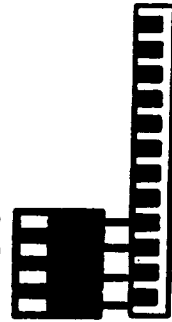


X-Y drive



Reagent cassette

ACGT



Microtiter plate

Lens array



CCD-camera

Positioning

Inkjet delivery

Mixing and thermostat

Light collection

Detection

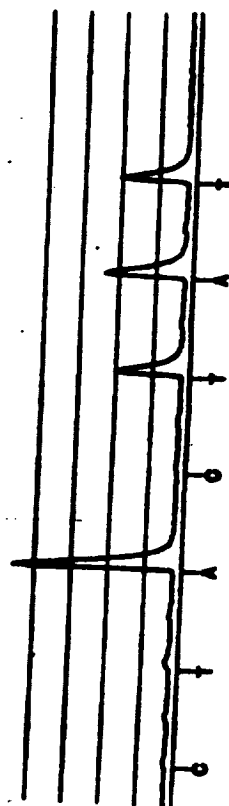
—A/TATA—
↑



a



b



c



SCIENCE AND TECHNOLOGY CENTER